PRODUCTION OF DOCUMENTS

Technical Field

This invention relates to the production of documents. The invention also
relates to the distribution of such reading material.

Background

Commercial computer networks, such as the Internet, have been used as a
means of facilitating ordering of books and other reading material by consumers. This
is typically achieved by presenting a web site-based user interface to consumers to
allow them to order reading material such as books. One example of this is the website
Amazon.com. However, the reading material that can be purchased by users of these
systems are the same as the offering made by a traditional book store. That is, each
item of reading material is usually offered in only one format. Further, users must wait
whilst the reading material they ordered is retrieved from a warehouse and shipped to
them.

The distribution of electronic documents is generally known and is described,
for example, in International Publication No. WO 00/72235 A1 (Silverbrook Research
Pty Ltd, 30 November 2000). Silverbrook describes text being formatted in the
Extendable Mark-up Language (XML) using the Extensible Stylesheet Language
(XSL). However, Silverbrook enables only a single user choice for formatting, namely
larger presentation.

Disclosure of the Invention

The invention discloses a system for producing a document comprising: a
repository for storing documents in a marked-up form according to one or more mark-
up schemas adapted to make explicit the structural information contained in a
document; a document format store for storing formats; and a document production
processor for generating a user-requested document from said marked-up form using a
user-selected format, said generated document retaining said implicit structural
information.

The invention further discloses a method of producing a document comprising

the steps of: marking-up a document according to a schema that makes explicit the structural information contained in said document; applying a user selected or created format to said marked-up document; and generating a user-requested document using a user-selected format, said generated document retaining said implicit structural

5   information.

The invention yet further discloses a system for producing and distributing a document comprising: a server site including a repository for storing documents in a marked-up form according to one or more mark-up schemas adapted to make explicit

10   the structural information contained in a document, a document format store for storing formats, and a document production processor for generating a user-requested document from said marked-up form using a user-selected format, the generated document retaining said implicit structural information; a network to which said server site is in communication; and a printing site to which said user requested document is

15   sent via said network to be printed.

The invention yet further discloses a method for distributing documents comprising the steps of: marking-up documents according to a schema that makes explicit the structural information contained in said document; receiving a customer

20   order for a said document over an electronic network, said order including formatting information; applying a user-selected format containing said formatting information to said marked-up document; generating a user-requested formatted document in electronic form using said format, the generated document retaining said implicit structural information; and transmitting said electronic document over said network.

25

Brief Description of the Drawings
In the drawings:

    . Fig 1 is a schematic view of an embodiment of a system for distributing

30   reading material according to the present invention; and

    Fig 2 is an example of a document marked up to an XML schema and stored in the system of Fig 1;

35   Fig 3 is an example of an XML document used to define a format in the system of Fig 1;

Fig 4 is an example of an XSL style sheet used in the system of Fig 1;

Fig 5 is an example of an XSL:FO document used in the system of Fig 1;

Fig 6 is an example of a PDF document output by the system of Fig 1;

Fig 7 is an example of a passage of text;

Figs 8 to 12 depict the text of Fig 7 with further alternative formats applied.

Figs 13 to 28 illustrate examples of text with formats applied.

Definitions

In this specification the following words have the following meanings:

Document - is intended to mean any reading material in hard copy or electronic form and includes books, pamphlets, brochures, reports, bank statements and other written material.

Format - is used to describe the general physical appearance of written material, including such things as type face, type size and margins.

Detailed Description and Best Mode

Referring to **Fig 1**, a system **5** for distributing reading material is shown in schematic form. The system **5** includes a network **6**, a client site **7** and a server site **10**. The server site **10** is represented schematically as a collection of software functions running on a suitably configured computing system which is connected to the network **6**, typically the Internet.

The server site **10** includes an interactive web site **12** which is presented to users (ie. clients), and allows users to request documents in a desired format.

A load/markup process **14** allows the upload and mark-up of documents to conform with a pre-defined set of rules, that preferably is a XML schema. The schema is constructed in such a manner as to facilitate automated publishing. Thus an

advantage of the schemas as described is that textual structural information is retained such that a coherent copy can be produced. For example, the line breaks in a poem are vital to its integrity as a document. The application of a schema makes the structural information implicit in a document become 'explicit'. There is a schema for each 'class' of document. A non-exhaustive list of classes includes: novel, technical text, engineering text, history text, and so on.

Each schema specifies 'major' structural information/elements of a document, which are elements that do not contain text directly. For a novel this typically can include: Book, Front matter, TOC, Preface, Introduction, Body, Chapter, Section, Sub-section, End matter, Index, and so on. Each schema further specifies 'minor' structural information/elements that contain text and, for example, emphasis. This typically includes: Para, Number para, and Special para. The purpose of the 'Special para' element is to avoid the need for excessive elements. A specific example of a 'Special para' is a poem, made up of a series of 'lines' (including blank lines), with attributes to handle justification and presentation. In other words, in most schemas, a 'line' is the highest level of precision contemplated in the mark-up schema, apart from words and characters required for special formatting (see below).

All 'minor' structural elements are required to flow into the rendition, and in that sense the flow is an immutable rule that can not be affected by the user formatting. The granularity of 'minor' structural elements can be as fine as individual words or characters, which would allow control over formatting down to the word or character level (as will be discussed below).

An example of a document marked-up to such a schema is shown in **Fig 2**. In this example, a 'minor' structural element is each of the paragraph tags <para number = " " > enclosing text. Documents are marked-up to this schema either by users or by the server site **10**, and stored in a repository **16** after being validated (eg. parsed). The repository could take many forms, including LAN-connected computers or multiple database servers connected over the Internet.

When a user orders a document for production, in addition to identifying the document, the user must specify or choose the format in which the document is to be produced. The web server **12** which allows a user to choose from a range of existing formats (ie. stored formats **22**). Alternatively, the user can prepare and select their own

format using format builder **20**, or take an existing format and change it. Formats are stored as XML documents (or in a database) where each format parameter (such as font type, font size etc) has an XML tag and attributes (or named value pair) that allow a style sheet builder **25** to recreate a style sheet that will generate the formatted document.

5   An example of an XML document used to define a format shown in **Fig 3**. Formats are discussed in greater detail below.

The XML documents embody a set of rules that encompass what is needed to process a document to completion. The rules have replaceable parameters that can be

10   chosen or modified by the user.

The format builder **20** therefore allows the user to specify desired format parameters based on personal requirements. The user can specify one or more of the following format parameters:

15   • page size

• margins

• fonts (including special fonts and word shaping)

• leading (including line shaping and making special effects in spaces between lines)

• effects

20   • colours

• spacing

• shading

• justification

25   The rules operate to transform the information made explicit in the stored marked-up documents back to implicit information embodied in the output document.

The format tester module **18** is an optional module that helps a user select the best format for that person as an individual. The format tester **18** operates on rules that

30   are based on knowledge of reading disabilities and formats that assist those reading disabilities.

The selection purchase process **24** allows a user to select a document they wish

to be produced from those stored in the repository **16**. The web server **12** enables a user to make a selection from one of a plurality of stored formats or to create their own format.

5   After the user has selected their required format, and the document they wish to be produced, the document is produced in the selected format by operation of the styles sheet builder **25**, an XSLT processor **29** and an XSL:FO processor **32**. The Style Sheet Builder **25** uses the XML file defining the format selected by the user to create an XSL:FO style sheet **27**, for example as shown in **Fig. 4**. This style sheet is then applied
10   by the XSLT processor **29** to the XML document which corresponds to the document required by the user from the repository **16** to produce an XSL:FO file **31**, for example as shown in **Fig. 5**. The explicit flow information in the XML document captured in the mark-up cannot be modified by this process. When in final form, the XSL:FO file **31** is processed by the XSL:FO processor **32** to produce the document in a form ready for
15   printing, in this case in PDF format **33**, as shown in **Fig. 6**.

         The server site **10** includes a print and delivery optimisation system **30** which arranges delivery of the produced document to the user based on order information provided by the user. The order information is a combination of known facts about the
20   user which are associated with their user ID such as their default delivery address and preferred payment method, in combination with any special requirements they have included in the particular order they are making. The order information also includes details of the selection of desired document and format that the user made above.

25   The delivery and optimisation system **30** may deliver the produced PDF file to a user by sending it to their e-mail address.

         The delivery system may have built-in file compression. The client site **7** may not embody the ultimate user/purchaser, rather can be a printery at a physical location
30   close to the reader which prints, binds and dispatches a hard copy document to the user. The printery can be selected for proximity, for lowest printing cost, for lowest printing and delivery cost, or for speed of delivery depending upon the requirements of the user.

         The production cost of the produced document is determined in part by the
35   format previously selected by the user. For instance, users with good eyesight can have books printed out in a small font and thereby require less paper. This lowers the

production cost. Users requiring a large font will need more pages printed so the produced document will cost more. The printing cost is also determined in part by the location of the printery. For instance, printing at a location near to the user will minimise the transport costs in transporting the printed document to the user's delivery

5   address. The printing cost is also determined in part by the country in which printing is carried out. For instance, printing in Mexico is much cheaper than the United States.

The financial systems module 26 collect any payments required from users as a result of use of the system 5.

10

Reader Defined Variable Format Patterns

As described above, the system 5 allows a user to define variable patterns within the format of a book or other document, and apply these format rules automatically to the whole document. Creating a book where every page is visually

15   different is an aid to visual memory. The variable format patterns of the book include:

• Random patterns – for example, every page is formatted with different margins or paragraph margins which are determined by random numbers

• Content based patterns – for example, every mathematical formula is printed in a particular way.

20   • Regular patterns – for example every page has a different watermark on it.

• A combination of the above

The variable format patterns include variations in the following parameters:

• Variable paragraph shapes

25   • Variable paragraph line spacing within and between paragraphs

• underlining with variable thickness and coloured lines

• creation of patterns of words in paragraphs to make a paragraph visually memorable, using fonts, colours, type sizes (one pattern in one paragraph and another in another paragraph)

30   • creation of patterns within paragraphs using the same technique (ie. one paragraph is in one format, the next paragraph is in another format etc)

• different watermarks

• varying page margin sizes

- placement of distinctive and possibly unrelated illustrations on a page

Using the system 5, a user could therefore obtain a textbook in two volumes – one for the text and the other for the diagrams, tables, footnotes, indices etc. Or the user could define large margins so that they could write notes in these margins. Further, a user could apply random page formats. Yet further, a user could insert blank pages at appropriate locations.

These formats can be applied to produce documents intended for electronic use or printing in hard copy documents and is not limited to the production of books.

Special Formats

Some people have trouble comprehending reading material. This can be for a number of reasons including problems with vision, eye control, discrimination of individual images, recognition and conceptualisation of images into meaningful words, and processing of meaningful word concepts into meaningful sentences.

Problems with vision may include lens problems involving focus such as astigmatism, long-sightedness, short sightedness and other lens problems, retinal problems such as inability to read in normal light conditions, colour contrast issues, blind spots, and nerve problems connecting the eye to the brain.

Problems with eye control include the inability of the eye to follow words sequentially in a line of text in the correct direction.

Problems with recognition and conceptualisation of meaningful words include the inability to differentiate between the image of a character and the mirror image of the character (eg. "b" and "d") or the same character after rotation (eg. "d" and "p"). They also include transposition of characters in a word, reading the whole or part of a word backwards as is common in dyslexia, and reading words in a different order to the order in which they are written. Another kind of problem is that people may not know what specific words mean.

Special formats are formats specifically designed to help people better discriminate characters they have difficulty in discriminating, and to provide additional information in the form of visual patterns that will assist readers mentally to process the

words and characters in the right order.

The formats that are applied to the document may cause it to be changed with respect to character height, character width, font colour, background colour, charac ter density, margins sizes, use of an optically corrected font, use of a shaded font, line length, line spacing, and separators between lines of text. A combination of the above may be used.

There are several different kinds of formats which can be applied. One set of formats aid discrimination between the characters or symbols presented to users. An example is to make the character "b" and the character "d" look so different that the reader can distinguish them. Another example is to format the text in a size, colour and font so that a person with visual impairment can read it.

Another example of a format involves the creation of a pattern in the characters and words to give additional information to the reader so that the reader can better interpret order of the characters and words. An example is to print text with words in alternating colours or "shape" so that words start in a small font and finish in a larger font, or vice versa.

Yet another example of a format is to add colour to words of a particular grammatical type, such and a noun in red and a verb in blue.

Yet another example of a format is to add additional information into the text, such as words in another language or pictograms.

Formats may also affect the spacing between lines. Lines of various thickness and shapes can be inserted between lines of text to help readers order characters within words, words within lines and lines within paragraphs.

The person may read the formatted document when it is in either hard copy or electronic form. With electronic materials, the format can be dynamic. An example is highlighting of words in a particular order for a particular time so that the reader's eye is taken along the line of the text in the right order, and without the eye jumping to the next line.

An example of an optically corrected font is tall thin characters for a person whose astigmatism elongates horizontally and contracts vertically.

Other formatting rules that may be applied include changes to the alignment of the characters at the bottom of a line to align them with the middle or top of a line.

Referring to **Fig 7**, an example passage of text is shown.

Referring to **Fig 8**, the text of **Fig 7** is shown with increased left and right margins.

Referring to **Figs 9, 10** and **11**, the text of **Fig 7** is shown with three different paragraph shapes. These are achieved by varying line length within the paragraph and justifying the text of the paragraph either to the left or the right.

Referring to **Fig 12**, the text of **Fig 7** is shown with increased line spacing.

Referring to **Fig 13** a line of text is shown formatted to give it a shape. The character height diminishes towards the middle of the line of text.

Referring to **Fig 14**, a line of text is shown formatted to give it a pattern. Alternate characters along the line are formatted in bold type.

Referring to **Fig 15**, a line of text is shown five times, each formatted to give a pattern. The pattern repeats in groups of two words. Every second word along the line is formatted in the same manner.

Referring to **Fig 16**, a line of text is shown formatted to give a pattern. The pattern repeats in groups of three words. Every third word along is formatted in the same manner.

Referring to **Fig 17**, a paragraph of text is shown formatted to give a pattern. The pattern repeats in groups of three lines of text. Every third line is formatted in the same manner.

Referring to **Fig 18** a line of text is shown formatted to give a pattern. The

beginning and end character of each word are formatted differently to the remaining characters of the words.

Referring to **Fig 19** a line of text is formatted to give a pattern. The pattern repeats in groups of two letters. Alternate letters have different character heights.

Referring to **Fig 20**, a paragraph of text is shown with lines of constant thickness inserted between each line of text.

Referring to **Fig 21**, the paragraph of text of **Fig 20** is shown with lines of tapered thickness with the taper extending in alternating directions for each alternate line of text.

Referring to **Fig 22**, three lines of text are shown each formatted to give a different pattern. Alternating groups of words are underscored with lines of varying thickness.

Referring to **Fig 23**, two lines of text are shown. In each line, identifying marks are associated with the characters "b" and "d" to assist a reader to distinguish between these characters.

Referring to **Fig 24**, there are two lines, one of words and the other of incrementing numbers situated under the middle of the words. These numbers can assist a reader to keep the words in sequence.

Referring to **Fig 25**, there is a line of words and a line of dots below the line of words. The first word has a single dot below it, the second word has two dots below it, and the third word, three. This pattern then repeats itself. The dots allow a user to sequence the words in the right order by providing more visual information about the order of the words.

Referring to **Fig 26**, a number 123,456,789 is shown in a format where the each three numbers are separated by commas. The second number is bigger than the first and the third number is bigger than the second. This gives a user visual information about the order in which the numbers occur and assists readers in keeping the numbers in the right order.

Referring to **Fig 27**, there are two lines, one of words and one of a pattern of alternating symbols used to distinguish between suits in a pack of cards. The position of the symbol approximately in the middle of the word sets up a visual pattern that

5   allows users to locate the words in the right order.

Referring to **Fig 28**, two triangles are situated below the letter "q", pointing left, and the letter "p", pointing right, providing a reader with more visual information to help distinguish between "p" and "q".

10

It is to be appreciated that various alterations or additions may be made to the parts previously described without departing from the spirit or ambit of the present invention.